

# Emotion Recognition in Psychotherapy: Impact of Therapist Level of Experience and Emotional Awareness



**Paulo P. P. Machado**

*Universidade do Minho*



**Larry E. Beutler**

*University of California, Santa Barbara*



**Leslie S. Greenberg**

*York University*

Accurately identifying another person's emotional state is an ability that may be necessary for a psychotherapist to empathize with a patient and that may be required for obtaining valid and reliable psychotherapy process ratings in research. Accuracy of identifying emotions and of rating emotional intensity expressed by a patient was studied in a comparison of 36 experienced therapists and 36 undergraduate psychology students who intended to become psychotherapists. Representative segments of a psychotherapy session were presented in one of three ways to tease apart the relative importance of verbal and nonverbal cues in making accurate ratings. Accuracy was judged against ratings supplied by two experienced and prestigious clinicians based on the same therapy sample. Results indicated that although therapists were more accurate than nontherapists in identifying emotions, they did not differ in the accuracy of rating emotional intensity. Moreover, accuracy of ratings was found to be less reliant on verbal cues among psychotherapists than among nontherapists. Finally, levels of participants' personal awareness of their own emotions had a positive impact on the accuracy of identifying specific emotions but not on the accuracy of rating their intensity. © 1999 John Wiley & Sons, Inc. *J Clin Psychol* 55: 39-57, 1999.

---

This study was partially supported by grant of the Junta Nacional de Investigação Científica e Tecnológica/National Council for Scientific and Technological Research, Portugal (JNICT-BD/1112/90/ID), and an American Psychological Association Dissertation Research Award to the first author. Correspondence concerning this article should be addressed to Paulo P. P. Machado, Universidade do Minho, Departamento de Psicologia., Campus de Gualtar, P-4710 Braga, Portugal.

Although it is generally agreed that making accurate judgments of another person's emotional state enhances the effectiveness of communication, the ability to accurately perceive other people's emotions is especially important in psychotherapy and psychotherapy research. Therapeutic empathy requires that a psychotherapist be able to recognize both the quality and intensity of a patient's emotional experience (Greenberg & Goldman, 1988; Greenberg & Safran, 1987, 1989; Ivey, 1983), and psychotherapy researchers must be able to identify objectively the emotions expressed in psychotherapy samples to accurately determine the role that emotional expression plays in psychotherapeutic improvement and change (Greenberg, Rice, & Elliott, 1993).

Considerable research has been devoted to identifying the factors that enhance and inhibit accurate emotional recognition. Much of this research has taken place in contexts other than psychotherapy, however, and frequently the dimensions identified lack relevance to this domain of communication. The current study was designed to (a) evaluate the degree to which some of the variables that have been identified may be generalized to the psychotherapy context and (b) identify how variables drawn from three separate lines of research interact to affect emotional recognition in psychotherapy.

The first body of research from which this study drew has been committed to uncovering the relative importance of verbal and nonverbal interpersonal cues in human communication. Although most of this research has been conducted outside of the psychotherapeutic context, it has much relevance for understanding psychotherapeutic processes. This research confirms both that nonlanguage vocalization and postural cues account for the preponderance of communication accuracy and that self-reports are frequently inaccurate indices of emotions about which one has conflict (Burgoon & Ruffner, 1978). It indicates, for example, that denied emotions often reveal themselves to be present through nonverbal behaviors over which individuals may have relatively less control than they do over verbal reports. Emotional recognition is dependent on information received through both verbal and nonverbal channels. Individuals seem to be able to recognize emotions with a fair level of accuracy, but this accuracy may decline when information is limited to either verbal or nonverbal channels of communication, such as facial expressions and vocal cues alone or when different channels provide contradictory information (Ekman, 1982, 1992, 1993; Ekman & Friesen, 1974; Scherer, 1981, 1986). Not only does this suggest the possibility of teaching psychotherapists to attend to nonverbal behaviors, but it calls into question the procedure of using either written transcripts of psychotherapy or patient self-reports as measures of emotional states in psychotherapy research.

The second line of research has been devoted to training psychotherapists to recognize and respond empathically to patient emotions. This research has sought to determine if the accuracy of emotional recognition and of empathic responding can be increased by teaching therapists and counselors to attend to nonverbalized information (Ivey, 1971, 1983). Although such specific and focused training has proven to increase the accuracy with which therapists can respond to patients' emotional states, its relevance to conventional training of psychotherapists is uncertain. Moreover, much of the research on this topic has been confined to analogue patients and therapy sessions, calling into question the justification of generalizations to clinical material.

Even when research on emotional recognition does include professionals who are conventionally trained and experienced (e.g., Johnson, Emde, Scherer, & Klinnert, 1986), it fails to compare their accuracy to individuals who are inexperienced and untrained. But even if the effects of training were adequately addressed, the question of generalization would not be solved. Typically the cues used to convey emotional states in such research are provided by actors who present preset verbal and nonverbal messages (e.g., Johnson et al., 1986). Paradoxically, this methodology has a built-in bias against recognizing

authentic emotional expressions, a fatal one if it is indeed true that deception is conveyed by subtle nonverbal cues (Ekman, 1992). Such research practices may yield results that do not represent the authentic display of conflicted emotions in naturalistic settings and psychotherapy practice (Wagner, MacDonald, & Manstead, 1986).

The third area of research of relevance to the current study has been that which has focused on aspects of the rater/therapist that might facilitate or inhibit accurate emotional identification. The level of clinical experience of the rater or judge has been one such dimension and personal sensitivity to one's own emotional experience has been another area of particular attention in the search for mediators of accuracy in recognizing emotions. Several studies (e.g., Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979) suggest that clinicians are more sensitive to nonverbal communication cues than teachers and business executives but, surprisingly, are somewhat less accurate than graduate students and actors. Indeed, a comparison of M.A. candidates, Ph.D. students, Ph.D. candidates, and clinical psychology faculty revealed a negative relationship between academic achievement and indicators of emotional sensitivity. Unfortunately, the level and amount of clinical training and experience among these groups was not assessed.

It is possible that insensitivity to emotional cues among formally trained individuals accounts for the low relationship between clinical experience and training, on one hand, and clinical effectiveness, on the other (Beutler, Machado, & Neufeldt, 1994). However, a number of authors (e.g., Greenberg & Safran, 1987; Mahoney, 1991) believe that lack of sensitivity to one's own emotions is a more likely candidate for understanding the variations that exist among clinicians in both accuracy of identifying others' emotions and clinical effectiveness. This hypothesis has yet to be systematically tested in clinical settings and relationships.

The current project was designed to determine if amount of experience and training as a psychotherapist enhances the accuracy of judging both verbal and nonverbal indicators of emotions among clients who are participating in real therapy sessions. It was also designed to determine if levels of personal emotional awareness mediate between either experience or type of stimulus material and accuracy of ratings.

Because the contemporary methods of assessing emotional quality and intensity rely on either patient self-report or therapy transcripts and because these materials exclude important aspects of emotional qualities, we found it necessary to first develop a measure that includes language, vocal, and nonverbal cues of emotional quality and intensity. Thus, the current project consisted of two separate, but related, studies. The first study was designed to evaluate if emotions occurring in a therapy session could be assessed and discriminated reliably. The second study was designed to evaluate the effect of several variables on the accuracy of ratings among participants who varied in training and experience background.

The second study postulated the presence of interactions among rater sensitivity to personal emotions, experience, and type of stimulus (language versus nonlanguage versus both) in determining accuracy of ratings. Specifically, we anticipated that accuracy among those with therapy experience would be less reliant on verbal content and more reliant on personal or self-sensitivity and on nonverbal material than those who do not have experience as psychotherapists.

## **STUDY I: DEVELOPMENT OF THE CLIENT EMOTIONAL AROUSAL SCALE**

Most measures of patient experience in psychotherapy ignore the role of nonverbal expressions. Others fail to distinguish among the widely varied emotional qualities that characterize people's interpersonal communication. For example, the widely used Experiencing Scale (Klein, Mathieu, Gendlin, & Kiesler, 1970) has proven to be helpful in identifying

the level of emotional intensity, but it does not differentiate among the types of emotions that may be represented in psychotherapy. The authors explicitly assumed that accuracy of rating the qualitative aspects of emotions was unimportant. Similarly, the Strength of Feeling Scale (Mahrer, Stalikas, Boissoneault, Trainor, & Pilloud, 1990), another measure that was designed to assess emotional experience, also omitted ratings of particular emotions. In contrast, the Affect Rating Scale (Koegel & Egel, 1979; Schreibman, Kaneko, & Koegel, 1991) defined the nature of positive emotions but is limited because of its focus on children and their parents in behavior therapy.

The Emotional Arousal Scale (EAS; Daldrup, Beutler, Engle, & Greenberg, 1988) is an exception to the rule of nonspecificity. However, as originally conceptualized, it was designed to assess the presence of anger and its role in psychotherapy. Accordingly, the presence and intensity of anger was rated. Because it could be applied equally well to verbal and nonverbal material and because preliminary work with it was promising, it offered a basis for expanding the focus to include six primary emotions and their intensities.

### **Method**

The objective of the Client's Emotional Arousal Scale was to provide a means for independent raters to identify the primary emotion(s) presented by the client in a therapeutic session or segment, as well as to derive an index of its (their) intensity. Scale development began by revising the items of the original EAS (cf. Daldrup et al., 1988, pp. 26–27) by expanding and extending the list of emotional experiences that were relevant to psychotherapy. Wherever the original scale identified anger as a rating target we added love, fear, joy, surprise, and sadness. In rewriting the items, we adopted three guiding principles: (a) categories of emotions should be nominal and mutually exclusive, (b) categories should represent the entire range of basic or primary emotions presented by clients in psychotherapy, and (c) the scale should provide clear ratings both of emotional quality and intensity.

The resulting, revised scale was divided into two parts. The first part of the scale was designed to assess emotional quality, or which primary emotion the client was expressing. Although these emotions are complex functional wholes, including cognitive appraisals, physiological responses, action tendencies, subjective feelings, expressions, and instrumental behaviors, they fit into families within which all members share some degree of resemblance but no universal set of features (Fisher, Shaver, & Carnochan, 1990). This first part listed six basic emotional experiences: love, joy, surprise, anger, sadness, and fear. Each of these emotional labels was presented with a list of emotional experiences that were related with each of the primary emotions. Both the labels of the prevalent emotions and the names of the associated emotional experiences were derived from Shaver, Schwartz, Kirson, and O'Connor's (1987) research and hierarchical model of emotion categories. The labels represent basic categories and the associated terms represent subordinate categories. This hierarchical model was built on research concerning American adults' use of emotional terms and was consistent with the findings for European and Chinese adults (cf. Fisher et al., 1990).

The second part of the scale was designed to assess clients' levels of emotional intensity and required the judge to make two ratings. For this part, we retained the previously established method for assessing the intensity of the emotion. One rating was derived for the modal and the other for the peak intensity of emotional arousal presented. Modal intensity represents an average expressed intensity and addresses the question, How intense was the emotional experience over its course? Peak intensity refers to the

maximum level of emotional arousal and addresses the question, How intense was the most intense moment of the emotional experience? (cf. Fridja, Ortony, Sonnemans, & Clore, 1991). This part of the scale was an adaptation of Daldrup et al.'s (1988) Emotional Arousal Scale intensity item. Wording was modified to accommodate the expanded list of emotions.

### *Assessment of CEAS Psychometric Properties*

The data set of this study consisted of 75 segments of group therapy interactions from the Arizona Treatment of Depression Study reported elsewhere (cf. Beutler et al., 1991). The therapists were four experienced (five or more years) psychologists (Ph.D.'s) selected from among eight professional-level candidates originally trained through a series of intensive workshops and practice exercises in the directive procedures of two different models of treatment. The patients were 63 moderately depressed outpatients selected from an initial pool of 376. The patients' age averaged 46.6 years ( $SD = 13.0$ ). All had (a) a DSM-III-R diagnosis of Major Depressive Disorder (MDD) and (b) a clinically significant score of 14 or more on the 17-item version of the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1967). This data set was chosen because it included a broad and contrasting array of psychotherapeutic treatments representing different approaches to emotional expression by clients in the therapeutic session; to ensure predictive and discriminant validity, the CEAS-R was tested to ensure that different patterns of arousal were represented in these treatments.

### *Procedure*

*Data Preparation.* Twenty-two segments of psychotherapy, representing two contrasting treatments, cognitive and gestalt therapies, were randomly selected from initial, middle, and end-of-treatment group therapy sessions. Each segment represented one client's full therapeutic work turn, and the group members' reaction (feedback) to that work. These segment units represented therapeutic episodes (Elliot, 1991): series of conversational sequences characterized by a common task or topic. The beginning of a conversational sequence was identified when a therapist initiated a response from the client about a topic and ended when another client intervened or the topic changed. The first client in each segment was the target of the therapist's intervention. The segment also included other participants' feedback to the client that followed the patient-therapist interaction. The length of these segments varied from 18 to 47 minutes ( $M = 25.8$  min.,  $SD = 7.72$ ), totaling 569 minutes of therapy sessions. The later ones varied in length with the amount of feedback that participants provided to the working client. In these segments 75 individual clients' interventions (working client or feedback) in the group were rated.

*Raters.* Two advanced doctoral students in a scientist-professional psychology program, who were blind to the purposes of this study, were recruited through an announcement sent to all students in the graduate program. The judges included one woman and one man who were willing to participate in the training. Raters were pretrained in the use of this scale in a two-hour workshop conducted by an independent researcher who was not involved in the development of the scale, but who had been previously briefed by the author on the use of the scale. The purpose of this workshop was to train raters in the use of several psychotherapy process measures; the amount of time spent to train in the use of

the CEAS was 30 minutes. Training consisted only in making raters familiar with the CEAS and its use.

**Rating procedures.** Judges were instructed to identify the quality of the emotional experience presented by the client and rate its intensity. While identifying emotional quality, raters were asked to choose from a list of six primary emotions (love, joy, surprise, anger, sadness, and fear) which ones were present during the rated segment. If more than one emotion was observed, they were to rank order the prevalence of each one. For the emotion rated as most prevalent, judges were asked to provide ratings of emotional intensity. First, the mode rating indicated the judge's evaluation of the client's average or usual emotional intensity during a given segment. Second, the peak rating corresponded to the judge's evaluation of the client's most intense expression of emotional arousal during the same segment.

## Results

Both proportion of rater agreement and Cohen's  $k$  were computed for the 75 ratings of the two judges regarding the quality of the emotional experience presented by the client during a specific interaction. The overall proportion of agreement for the expression of the prevalent emotion was  $po = .87$ . When this proportion was corrected for chance agreements utilizing Cohen's  $k$  the results were  $k = .83$  for the complete scale. When proportions of agreement were computed for the presence or absence of one emotional experience at a time they varied from .81 to 1.00; kappa values ranged from .63 to .90, mean kappa = .81. Kappa values for each emotion were: love,  $k = .83$ ; joy,  $k = .87$ ; anger,  $k = .90$ ; sadness,  $k = .78$ ; fear,  $k = .63$ . Surprise was never identified by any of the judges, which makes the proportion of agreement equal to 1 and the  $k$  not interpretable. Kappa values above .75 were taken to indicate strong agreement between judges and the presence of consensual validity (Fleiss, 1981).

The ratings of mode and peak intensity of the most prevalent emotion on a 7-point, anchored rating scale ranged from 1 (*Client does not express any feelings*) to 7 (*Arousal is full and intense . . .*). These ratings, produced by each judge, were compared by first assessing the difference between the mean ratings of each judge and then using a Pearson's product-moment correlation coefficient. The means and standard deviations of the ratings of each judge for the mode and peak intensity of the prevalent emotional experience presented by the client in each segment were for Judge A: Mode-Mean = 2.4, Standard Deviation = 1.14, Peak-Mean = 3.1, Standard Deviation = 1.21; and for Judge B: Mode-Mean = 2.7, Standard Deviation = 1.21, Peak-Mean = 3.4, Standard Deviation = 1.53.

The mean ratings of mode and peak intensity by the two independent judges were not significantly different (Mode:  $t(148) = -1.52, p > .05$ ; Peak:  $t(148) = -1.19, p > .05$ ). The interrater correlation coefficient for the ratings of mode arousal intensity was  $r = .66$  and for the peak arousal intensity  $r = .71$ .

A separate validity analysis was undertaken on the CEAS-R (Rosner, 1996) using samples from the same data set. This analysis compared the level of arousal and types of emotions presented by manualized versions of gestalt and cognitive therapies. The results confirmed the sensitivity of the CEAS-R to different emotional states and demonstrated the predictive and discriminant validity of the scale. It revealed that the two therapies gave rise to significantly different patterns of emotions and different levels of emotional experience during the course of the two psychotherapies.



### **Conclusions**

We concluded that the psychometric properties of the Client Emotional Arousal Scale-Revised were adequate. We were able to establish reliability in terms of interjudge agreement levels for both emotional quality and intensity, and Rosner (1996) complemented these findings by establishing the predictive, discriminant validity of the measure. These results indicate that both quality and intensity of clients' emotional expression can be reliably identified and validly measured in therapeutic segments in a way that is similar to what therapists do in therapy sessions. Agreements among raters were high even after we corrected them for chance, indicating strong agreement between judges. When individual emotional experiences were analyzed, anger, joy, and love were the emotions in which judges reached the highest level of agreement. Although still at an acceptable level, ratings of fear didn't reach the same level of agreement as other emotional qualities. Surprise was never identified as present in any of the 75 segments. The judges' ratings of mode and peak level of emotional arousal intensity showed acceptable level of agreement. The intrajudge ratings of modal and peak intensity of emotional arousal were highly correlated, suggesting that these two ratings may represent a common underlying construct.

Summarizing, the Clients Emotional Arousal Scale provides a promising tool to rate and assess the quality and intensity of clients' emotional experience in the therapeutic context. The results also justify the initiation of Study 2 with the assurance that emotional experience can be reliably measured.

## **STUDY II: RECOGNITION OF EMOTIONS**

### **Method**

The second study had four objectives: (a) to determine if the quality of emotional states could be recognized during a psychotherapy session with an actual patient, (b) to determine the relative importance of verbal and nonverbal cues in making accurate ratings of quality and intensity, (c) to determine the degree to which participants' own levels of emotional awareness interfere with or enhance the accurate identification of types and levels of emotions in clients, and (d) to determine how rater personal awareness affected accuracy of ratings.

### **Design**

This study followed a  $2 \times 3$  factorial design in which the independent variables represented two levels of participants' experience (training) and three stimulus conditions varying in complexity of verbal and nonverbal emotional cues. In addition, the study incorporated one covariate (levels of emotional awareness) that was expected to interact with the other variables. Dependent variables included accuracy of identification of emotional quality (i.e., recognition of the prevalent emotion present) and emotional intensity (i.e., rating of modal and peak level of emotional intensity). A post-hoc analysis also was conducted to explore the subsidiary hypothesis that therapists from different theoretical orientations and models of training would show differential levels of accuracy of emotion recognition.

### **Stimuli**

Participants rated eight segments of one psychotherapy session carefully selected to represent a range of emotions and levels of emotional intensity. The original therapy session

represented the work of one client who had participated in the University of Arizona Depression Treatment Project (Beutler et al., 1991). The treatment used was a variant of gestalt therapy, Focused Expressive Psychotherapy (FEP; Daldrop et al., 1988). This treatment encourages the expression of intense and abreactive feelings. It was thus ideal for testing therapists' ability to recognize emotions and their intensities.

This study was designed to maximize generalization among raters (therapists and nontherapists). Thus, we conserved degrees of freedom by treating the therapy stimuli as a within-treatment variable, providing multiple segments from a single but representative therapy session in which the therapist encouraged the genuine expression of a range of emotional experiences. The use of multiple observations of a single participant followed the logic as an  $N = 1$  study, reflecting our conclusion that demonstrating that the emotions of an individual patient could be recognized is a more meaningful test of therapist sensitivity, clinically, than a demonstration that group averages could be affected.

To increase representativeness, the therapy session used as stimulus for this study was selected after a review of all the available sample work tapes of all clients that participated in the FEP groups ( $N = 22$ ). We selected eight segments lasting from 45 seconds to 1 1/2 minutes, totaling 10 minutes and 24 seconds. The segments started with either the therapist's or client's intervention and stopped when the client shifted topic, presented a notable emotional change (quality or intensity), or when the therapist redirected the client to other topics or activities. Long therapist speaking turns were deleted from the stimulus tape. The segments were arranged chronologically on a stimulus tape and corollary, verbatim transcripts were produced of the selected segments.

### ***Verbal and Nonverbal Stimuli***

Based on the segments described in the foregoing, three kinds of stimuli were produced to vary the balance of verbal and nonverbal cues. The first was a chronologically arranged ordering of the eight selected segments, combining both verbal and nonverbal cues. The second was a videotape of the same eight segments, but with the audio content filtered out to represent nonverbal stimuli. The third was a verbatim transcript of the previous videotaped segments to represent only verbal content cues. All the segments were presented in the same order on the three stimuli conditions.

The procedure used to produce the content-filtered videotape was similar to the methods described by Shoham-Salomon, Avner, and Neeman (1989). Content-filtered stimuli (cf. Rogers, Scherer, & Rosenthal, 1971) were produced by passing the audio channel of the videotape segments through two Allison 2BR low-pass filters set at 580 Hz. This procedure produced videotape segments in which frequencies above 580Hz were attenuated; as a result the emotional tone was preserved whereas the speech content was no longer identifiable. The settings for the frequency were chosen after the author listened to various possible settings and rated them on a subjective criterion of overall intelligibility. The setting was selected to ensure that the first author (PPM), who was familiar with the content of the segments, could not understand the words.

### ***Measures***

*The Clients Emotional Arousal Scale—Revised (CEAS-R).* The CEAS-R described in Study 1 was used. The CEAS-R required two ratings, one of the presence of emotional quality and another of intensity of emotional arousal.

*The Level of Emotional Awareness Scale (LEAS, Lane, 1991).* The LEAS was used to assess participants' personal level of emotional awareness. This scale consists of 20 scenes,



each describing in two to four sentences a situation involving interactions between two people. The scenarios were constructed to elicit four types of emotional experience: anger, fear, happiness, and sadness. One scenario was presented per page, followed by two questions: How would you feel? and How would the other person feel? Participants were required to write their responses on the remainder of each page. They were instructed to use as many or as few words as they needed. Responses were coded and received a separate score for each of the emotions described by the self and other. The lowest score (Level 0) is for "nonemotion responses where the word *feel* is used to describe a thought rather than a feeling." The highest score (Level 4) corresponded to "high emotional differentiation." Each participant thus received a separate score for the rated response of self and for the response attributed to others. The total score is obtained by adding the level for all 20 scenarios, thus ranging from 0 to 80. The total score is the sum of the two partial scores (Self and Others). The LEAS has shown adequate psychometric properties with high interrater reliability: interclass (20) = .84 (Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990).

*Dependent Variables: Accuracy in Rating Quality and Intensity.* It was necessary to establish an external criterion by which to define participants' accuracy in rating the nature of the emotional experience of the client during the stimuli therapy samples. Because it was impossible to contact the client we used experts' ratings as a standard against which participants ratings were compared. Although this procedure might sound somewhat artificial, it is not uncommon in psychotherapy supervision, for example, where the supervisor serves as an expert with no access to the client's actual experience. The experts were two recognized experts in the field of psychotherapy research and emotional processes in psychotherapy. They were coauthors of the FEP therapy manual, and each was a principal investigator of a program in psychotherapy research funded by the National Institute for Mental Health (LEB and LSG). None of them was the therapist portrayed or was familiar with the stimulus tape.

The experts were asked independently to view the unaltered videotape containing the eight segments of the selected therapy session. They were instructed to rate the emotional quality and then the emotional intensity expressed by the client after viewing each segment. Experts used the CEAS-R. Emotional quality ratings corresponded to one of the six primary emotions that were described on the CEAS-R form. Emotional intensity was the modal and peak intensity rating of the emotion using a 7-point, anchored Likert scale of the "most prevalent" emotion.

Although a consensus procedure was selected a priori for defining the criteria emotion, there was no need to use it because the experts' ratings were in complete agreement. Thus, the emotion that the experts consensually rated as being present in each segment was used as a standard against which to assess participants' accuracy of emotion recognition.

A different procedure was used to establish the standard for modal and peak emotional intensity. Because these ratings were not categorical, the standard for agreement was obtained by averaging the ratings of the two experts for each segment. For each rating of intensity, in those where the experts didn't agree, the standard became the average of the two ratings. Expert ratings never differed by more than one unit on the 7-point scale. The correlation coefficients for the experts' ratings were: mode,  $r = .57$ ; peak,  $r = .77$ .

### ***Participants***

Thirty-six upper-level undergraduates and 36 therapists participated as participants to represent two levels of experience and training but similar levels of interest in psycho-

therapy as a profession. Undergraduate participants were recruited from an introductory course on counseling psychology and were all volunteers. This group had some interest in the field and a basic knowledge of therapy procedures and purposes, and most anticipated entering a helping profession. Therapists were volunteers recruited through a general mailing to those in two counseling and clinical psychology graduate programs and from two psychotherapy research projects associated with them.

*Undergraduates.* Thirty-six undergraduate students from a large West Coast university in the United States were recruited as participants for this study. Inclusion criteria were: (a) no previous experience as a peer counselor or counselor, (b) university enrollment at the undergraduate level, and (c) no previous graduate work in psychology. The objective was to obtain a sample of participants with interest in counseling and psychotherapy who could be described as minimally trained. Students indicated that they intended to apply to a professional graduate program in psychology. Undergraduate participants were 30 women and 6 men whose ages ranged from 20 to 52 years old; the mean age was 25.17 years ( $SD = 8.11$ ).

*Therapists.* Therapist participants were either doctoral students, experienced psychologists, psychiatrists, or practicing master's-level therapists involved in one of two psychotherapy outcome research studies. Thirty-six therapists participated in the present study. Eighteen therapists were recruited from the Counseling/Clinical/School Psychology program (counseling and clinical psychology specialty areas) at a West Coast university in the United States and 18 were recruited from the Clinical Psychology program at a Canadian university. Therapist participants were 23 women and 13 men with ages ranging from 25 to 51 years old; the average age was 34.8 years ( $SD = 6.49$ ). Therapist participants were five clinical or counseling psychologists, one psychiatrist, 26 advanced doctoral students, and four master's-level therapists (i.e., MSW or MFCC). Advanced doctoral students were those in the last year or internship year of their graduate training who had completed all their practical training. Twelve of the 26 graduate students also participated as therapists in the aforementioned psychotherapy research projects. Therapists had an average of 5.6 years ( $SD = 4.09$ ) of experience as therapists. Most of the therapists had seen at least 10 clients in individual therapy.

The two samples were comparable in terms of most of the demographic characteristics. Undergraduate sex distribution was 30 women and 6 men; therapists were 23 women and 13 men. The majority of participants in both groups were Caucasians. In the undergraduate group 10 identified themselves as minorities, whereas in the therapist group, three did so. The only demographic variable on which the samples were significantly different was age; the therapists' age was significantly higher than the undergraduates' ( $t(70) = 5.55, p < .01$ ).

### ***Procedure***

Participants were randomly assigned to one of the experimental conditions. Participants were told that their task would be to evaluate the emotions expressed by the therapy client as presented either in a stimulus tape or on a transcript. All participants in the videotape or content-filtered condition were instructed on how to use the videotape recorder properly. They were then asked to familiarize themselves with the rating scales that they were about to use. Once ready, they were instructed to start the VCR and watch the first

segment. When the first segment was over they were instructed to pause the videotape and complete the ratings of emotional quality and intensity using the CEAS-R. They were to repeat this procedure for the remaining seven segments. Participants in the transcript condition followed a similar procedure adapted to the fact that they would be reading the transcript and basing their decisions on written material. All participants completed a personal demographic questionnaire and the Levels of Emotional Awareness Scale in a randomized order. The stimuli material and the questionnaires were presented in a randomized order for each participant to control for order effects.

## RESULTS

### *Preliminary Analysis*

As a preliminary analysis and in an effort to increase statistical power, we compared the therapists at the two sites on several indices. Within the therapists' group at Site A there were 10 women and 8 men, and at Site B there were 13 women and 5 men. These frequencies were not significantly different. Nonsignificant differences were also observed for the distribution of ethnicity, but the average age of therapists in Site B ( $M = 36.9$ ,  $SD = 6.32$ ) was slightly higher ( $t(34) = 2.10$ ,  $p < .05$ ) than the age of therapists in Site A ( $M = 32.6$ ,  $SD = 6.06$ ).

When the therapist sample was compared between sites on indices of experience and training, none of the analyzed variables showed significant differences. Therapists at both sites had similar mean years of practice as a therapist (Mean A = 5.1; Mean B = 6.1;  $t(34) = .77$ ,  $p > .05$ ), and both had similar numbers of trainees in supervision (Mean A = 5.8; Mean B = 3.4;  $t(34) = .89$ ,  $p > .05$ ). Finally, therapists in both sites had similar amounts of experience in terms of number of clients seen in individual therapy. Thus, for subsequent analyses, therapists at the two sites were collapsed into a single group.

### *Emotional Quality*

Accuracy of identifying the particular emotion expressed was indexed by the degree (proportion) to which participants agreed with the experts in identifying the primary emotion presented by the client in the eight segments. Each participant identified one or more of six primary emotions as being represented in each of the eight therapy segments sampled. If they rated two or more emotions as being present they were asked to rank order the prevalence of each one. To score accuracy, we scored only the one ranked as most prevalent. For example, if a participant had identified anger, sadness, and fear as present in a given segment and had ranked them in that order, we scored anger as the participant's rating of emotional quality. If this was the same as that given by the experts it was called accurate. If not it was called inaccurate.

For each segment, an agreement earned a score of 1, and a disagreement earned a score of 0. A percent of agreement score across the eight segments was computed for each participant by adding the number of agreements and dividing by eight. High scores represented an increasingly higher proportion of accuracy.

Because of the proportional nature of this dependent variable we used loglinear transformations of our raw proportional scores (Kirk, 1982). Means and standard deviations of the loglinear transformations of the proportion scores for each group and condition stimulus are presented on Table 1. These findings, when translated to proportions of agreement, reveal high levels of agreement with the experts ( $M = .76$ ,  $SD = .15$ ).

**Table 1.** Means and Standard Deviations of Accuracy for Each Group, Condition, and Group by Condition

Group	Undergraduates			Therapists			Total (row)		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Videotape	12	-.23	.16	12	-.26	.17	24	-.25	.16
Transcripts	12	-.32	.22	12	-.27	.23	24	-.30	.22
Content filtered	12	-.53	.34	12	-.21	.08	24	-.37	.29
Total (column)	36	-.36	.27	36	-.25	.17	72	-.30	.23

*Note.*—Row totals represent accuracy by condition and column totals represent accuracy by group.

A 2 (experience) by 3 (stimulus condition) analysis of variance (ANOVA) was undertaken to assess differences in level of accuracy in identifying emotional quality among groups, conditions, and interactions. The between-stimulus condition (i.e., kind of stimuli) effect was not significant ( $F(2,66) = 2.11, p > .05$ ). However, both between-group differences ( $F(1,66) = 5.30, p < .05$ ), and the group by stimulus condition interaction ( $F(2,66) = 4.42, p < .05$ ) were significant. Results revealed that accuracy was significantly higher for therapist than for undergraduate participants.

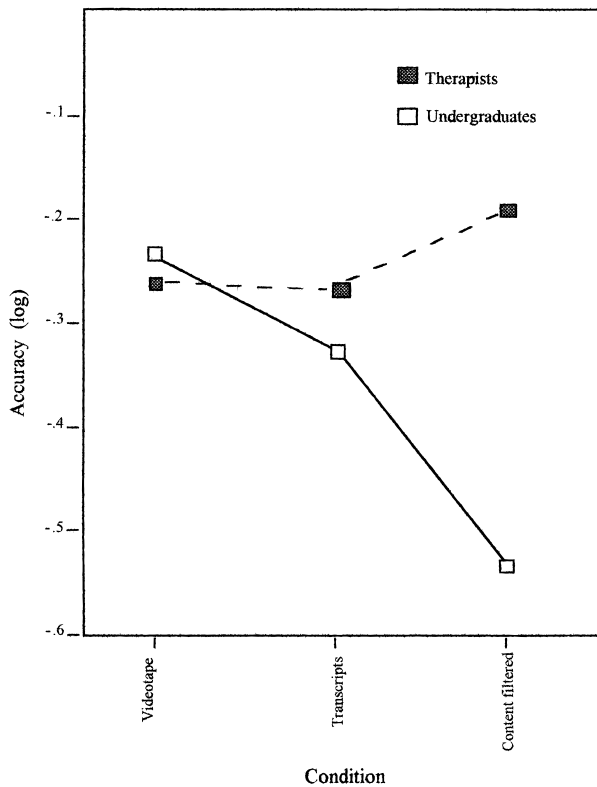
Figure 1 shows the group by stimuli level of accuracy in rating the emotional quality of the sample client; higher results indicate higher accuracy. Post-hoc analysis of the interaction effect revealed that the highest level of accuracy occurred among therapists rating content-filtered segments, whereas the lowest level of accuracy occurred among undergraduates when rating these same segments. This difference was statistically significant (Scheffé = .30,  $p < .05$ ). Undergraduates also reached significantly ( $p < .05$ ) higher levels of accuracy when rating videotapes than when rating content-filtered segments. All other differences were nonsignificant.

### ***Emotional Intensity***

In the second step of the analysis, the "agreement" scores were used to analyze emotional intensity data. Agreement was defined as the degree to which participants rated the level of emotional intensity, regardless of the specific emotion identified, at a level that agreed with the experts' mean estimates. For each participant, two separate deviation scores (mode and peak) were computed and were used as primary data for the subsequent analyses. Deviation scores represented the distance between each participant's rating and the experts' mean estimate. Higher scores represent higher deviations from the experts' estimate, and thus lower levels of agreement. The deviation score was computed by adding the squared differences of each score and the one of the experts for each segment, using the formula:

$$DRT = \sum_{i=1}^8 (R_i - R_{exp_i})^2$$

where DRT was the rating deviation score (either mode or peak) for each participant,  $R_i$  was the rating of each participant for a given segment, and  $R_{exp_i}$  was the expert's mean rating for the same segment. One score represented the modal rating and the other the peak rating. Modal rating indicated the participant's evaluation of the client average or usual emotional intensity during a given segment. The peak rating corresponded to the



**Figure 1.** Accuracy (Proportion of agreement—log) when therapists and undergraduates rated emotional quality in different stimuli conditions.

participant's evaluation of the client's most intense expression of emotional arousal during the same segment.

To protect experiment-wise alpha level, a 2 (experience) by 3 (stimulus condition) multivariate analysis of variance (MANOVA) was conducted with deviation scores for mode and peak as dependent variables. Results revealed that the group effect was not significant, Pillai's  $V = .06$ ,  $F(2,65) = 4.94$ ,  $p > .05$ . The stimuli effect was significant, Pillai's  $V = .26$ ,  $F(4,132) = 2.72$ ,  $p < .01$ . As the results were significant, two separate 2 (group) by 3 (condition) analyses of variance (ANOVAs) were conducted to assess differences between groups and conditions for each of the agreement dependent variables.

*Mode.* Table 2 presents means and standard deviations of the mode deviation scores representing levels of agreement in rating the modal level of emotional intensity in different groups and conditions. Deviation scores represent the distance between each participant's ratings and the experts' ratings. Higher scores thus represent lower levels of agreement.

There were no significant group differences (i.e., undergraduates vs. therapists) in level of agreement rating the mode level of emotional intensity ( $F(1,66) = 2.72$ ,  $p > .05$ ). However, the between-condition effect was significant ( $F(2,66) = 8.30$ ,  $p < .01$ ). The group by condition interactions were not significant ( $F(2,66) = .80$ ,  $p > .05$ ). Post-hoc Scheffé analysis of the condition effect showed that levels of agreement were signif-

**Table 2.** Means and Standard Deviations of Agreement (Mode Ratings) for Each Group, Condition, and Group by Condition

Group	Undergraduates			Therapists			Total (row)		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Videotape	12	8.92	3.97	12	9.17	7.60	24	9.04	5.93
Transcripts	12	18.00	11.2	12	13.58	6.19	24	15.79	9.15
Content filtered	12	10.00	8.45	12	5.66	3.11	24	7.83	6.61
Total (column)	36	12.31	9.17	36	9.47	6.64	72	10.89	8.07

*Note.*—Row totals represent agreement by condition and column totals represent agreement by group.

icantly higher ( $p < .01$ ) on the videotaped and content-filtered condition than on the transcripts condition.

*Peak.* Table 3 presents means and standard deviations of peak deviation scores representing levels of agreement rating the peak level of emotion intensity in different groups and conditions. Higher scores represented lower levels of agreement.

Like the first, or Mode, analysis, there was no significant group difference (i.e., undergraduates vs. therapists) in level of agreement rating the peak level of emotional intensity ( $F(1,66) = 3.84, p > .05$ ). The between-condition effect was significant ( $F(2,66) = 10.49, p < .01$ ), and the group by condition interaction was not significant ( $F(2,66) = .24, p > .05$ ). Post-hoc Scheffé analysis of the condition effect showed that levels of agreement were significantly higher ( $p < .01$ ) on the videotaped and content-filtered condition than on the transcripts condition.

Figure 2 presents the group by stimuli level of agreement in rating emotional intensity as expressed in deviation scores. To facilitate the graphic reading, all scores were multiplied by  $-1$ , so that higher scores represent higher levels of agreement with the experts judging levels of emotional intensity.

### Levels of Emotional Awareness

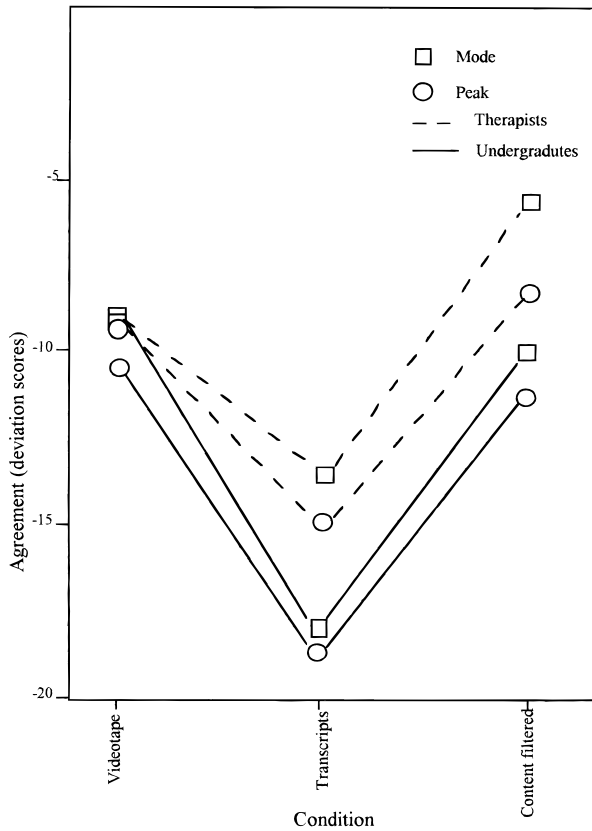
We had hypothesized that level of personal emotional awareness would either interfere with or enhance the accuracy and agreement. Analyses of covariance (ANCOVAs) were

**Table 3.** Means and Standard Deviations of Agreement (Peak Rating) for Each Group, Condition, and Group by Condition

Group	Undergraduates			Therapists			Total (row)		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Videotape	12	10.68	3.11	12	9.25	5.85	24	9.95	4.64
Transcripts	12	18.50	8.59	12	14.83	6.21	24	16.66	7.56
Content filtered	12	11.33	6.10	12	8.17	4.39	24	9.75	5.44
Total (column)	36	13.50	7.13	36	10.75	6.14	72	12.13	6.75

*Note.*—Row totals represent agreement by condition and column totals represent agreement by group.





**Figure 2.** Agreement level when therapists and undergraduates rated modal and peak level of emotional intensity in different stimuli conditions (higher scores represent higher agreement).

used to assess the effect of the Level of Emotional Awareness Scale and its interactions for the dependent variables of both accuracy and agreement. A pretest of the covariate was conducted to justify interpreting the interactions. The difference between group means on the scores of the LEAS-Self and LEAS-Other were not significant (LEAS-S,  $t(70) = 1.842$ ,  $p > .05$ ; LEAS-O,  $t(70) = 1.208$ ,  $p > .05$ ). The absence of significant effects indicated that the level of emotional awareness was not confounded with the group effect.

To test the hypothesis that levels of emotional awareness interacted with level of experience, an analysis of covariance (ANCOVA) was performed. The LEAS-Self score served as a covariate and accuracy of identification of emotional quality was used as a dependent variable. A significant effect of the covariate emerged ( $F(2,65) = 3.39$ ,  $p < .05$ ) in this analysis indicating that levels of emotional awareness alone accounted for increased accuracy once the effects of group and stimuli were partialled out. The effect initially observed for experience level washed out when the level of emotional awareness covariate was added ( $F(2,65) = 3.39$ ,  $p > .05$ ), suggesting that emotional awareness may be related to experience and experience effects were subsumed by the level of emotional awareness. However, experience by stimuli interaction effects were still significant ( $F(1,65) = 4.56$ ,  $p < .05$ ), indicating an effect beyond that of the therapist's level of emotional awareness. These results indicated that level of emotional awareness enhanced therapists' levels of accuracy in identifying emotional qualities.

Although our results didn't meet the criteria for a significant interaction between covariate and condition (Snow, 1991), the pattern of results indicates that participants with higher levels of emotional awareness were more accurate in identifying emotional quality than participants with low self-awareness. For both groups and for the three stimuli conditions LEAS-S scores were positively correlated with accuracy in identifying emotional experience. The correlation across groups and conditions between LEAS-S scores and agreement was  $r = .25$  ( $p < .05$ ).

The effects of the level of emotional awareness were not significant when analyses of covariance were performed with the agreement on ratings of emotional intensity as dependent measures (Mode,  $F(1,65) = .02$ ,  $p > .05$ ; Peak,  $F(1,65) = .04$ ,  $p > .05$ ). Finally, separate analyses of covariance (ANCOVAs) were performed to assess the impact of scores of the Level of Emotional Awareness Scale-Other on the dependent variables. The effect of the LEAS-O scores as a covariate was nonsignificant on all dependent measures (accuracy: proportion of agreement,  $F(1,65) = 3.13$ ,  $p > .05$ ; agreement: mode,  $F(1,65) = 0.00$ ,  $p > .05$ ; peak,  $F(1,65) = .05$ ,  $p > .05$ ).

## DISCUSSION

Collectively, the results of the two studies reported here support the differential and additive role of verbal and nonverbal cues in accurately identifying emotional qualities, the importance of therapist level of training and experience in determining how these cues are used, and the role of personal awareness in enhancing emotional recognition in others. Training as a therapist was associated with a heightened ability to accurately recognize the emotional experience of clients portrayed in the therapy segments independently of the stimuli conditions. There was no general or main effect of the stimulus condition on accuracy of emotional recognition, but accuracy was differentially responsive to different cues among trained and untrained therapists.

As we had hypothesized, when content was filtered out therapists still were able to accurately identify the quality of the client's emotional experience, whereas undergraduates' accuracy level declined. Overall, therapists were more accurate than their non-trained counterparts in identifying naturally occurring emotional experiences in the context of a therapy session. They seem to make optimal use of nonverbal cues to correctly identify the emotion presented by clients. Undergraduate raters, on the other hand, did not seem to be able to make sense of nonverbal cues when they were presented alone. They were able to obtain reasonably high levels of accuracy when they had access to verbal cues, however.

It is interesting to note that the content-filtered condition, when rated by therapists, produced the highest levels of accuracy, whereas the same condition when rated by undergraduates produced the lowest level of accuracy. In this psychotherapeutic context, non-trained individuals tended to be less able to use nonverbal cues than trained ones and thus were more reliant on the verbal contextual cues. Nevertheless, these contextual verbal cues might have been misleading and compromised the level of accuracy in identifying the primary emotions.

Although factors that are incidental to obtaining training (age, admission to graduate training, training itself, experience, intellectual or academic potential, etc.) cannot be ruled out in accounting for the differences noted between trained and untrained groups, the results seem to shed some light on the debate between two apparently antagonistic positions in emotion identification theory (see Scherer, 1986). Psychologists usually assume the existence of a direct covariation (covariation model) of emotional meaning with a

particular set of nonverbal cues independent of vocal content. One of the assumptions, for example, is that fundamental frequency ( $F_0$ , heard as pitch) is linearly related to the level of arousal. Linguists, on the other hand, defend (configuration model) the view that participants infer emotional meanings on the basis of the interaction between linguistic structure (content) and vocal cues, thus claiming that most emotion recognition studies (those that do not provide content cues) are unlikely to reveal the configurational cues that individuals use in differentiating subtle differences in expressed emotions. Previous findings (e.g., Scherer, Ladd, & Silverman, 1983) have suggested that the configurational model describes the communication of more cognitive affect states (e.g., doubt), whereas the covariation model seems more adequate for arousal-related emotions such as anger. The results of our study showed that trained individuals were able to infer emotional meaning in the absence of contextual cues (i.e., content-filtered condition), but untrained individuals were not able to infer emotional meaning in the absence of contextual cues, supporting to some extent the covariation model.

Both therapists and undergraduates were able to correctly identify the levels of emotional arousal independently of the emotional quality. However, when participants based their ratings on the verbatim transcripts of the therapy session, levels of agreement with the experts were significantly lower than in the videotape or content-filtered conditions. Not surprisingly, to correctly identify the level of emotional arousal participants needed to have access to nonverbal cues. Ratings based on content information alone tended to be misleading and inaccurate. These results tend to support the findings (cf. Scherer, 1981) that vocal cues (e.g., pitch) are linearly related with the degree of arousal and play a crucial role in the recognition of emotional arousal.

Finally, our hypothesis that levels of emotional awareness would interact with accuracy of identification of emotions and their intensity was partially supported. Levels of personal emotional awareness had a positive impact on the accuracy of identification of emotion, but not on the accuracy of rating intensity. The ability to focus on our own emotional process or to resonate to others' emotional experiences in interpersonal situations is likely to provide us with important information about others' experiences, enhancing our ability to recognize emotions.

In light of our results, we suggest that trained individuals are able to identify the basic categories in which emotional experiences are being organized (e.g., emotional families; see Fisher et al., 1990) by an individual in the absence of contextual cues. However, we do not know if these contextual cues are important in the identification of subordinate categories (Fisher et al., 1990) or more complex emotions such as pride, jealousy, or grief. Untrained individuals, however, were not able to infer emotional meaning from these subtle noncontextual cues. Thus, it appears that although nonverbal cues are a powerful component of interpersonal communication, its accurate decoding can be enhanced by specific training.

## REFERENCES

- Beutler, L.E., Engle, D., Mohr, D., Daldrup, R.J., Bergan, J., Meredith, K., & Merry, W. (1991). Predictors of differential response to cognitive, experiential, and self-directed psychotherapeutic procedures. *Journal of Consulting and Clinical Psychology*, 59, 333-340.
- Beutler, L.E., Machado, P.P., & Neufeldt, S. (1994). Therapist variables. In S. Garfield & J. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (pp. 229-269). New York: John Wiley & Sons, Inc.

- Burgoon, M., & Ruffner, M. (1978). *Human communication*. New York: Holt, Rinehart, & Winston.
- Daldrup, R.J., Beutler, L.E., Engle, D., & Greenberg, L.S. (1988). *Focused expressive therapy*. New York: Guilford.
- Ekman, P. (Ed.). (1982). *Emotion in the human face* (2nd ed.). New York: Cambridge University Press.
- Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science*, 3, 34–38.
- Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, 48, 384–392.
- Ekman, P., & Friesen, W.V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology*, 29, 288–298.
- Elliot, R. (1991). Five dimensions of therapy process. *Psychotherapy Research*, 1, 92–103.
- Fisher, K.W., Shaver, P.R., & Carnochan, P. (1990). How emotions develop and how they organize development. *Cognition and Emotion*, 4, 81–127.
- Fleiss, J.L. (1981). *Statistical methods for rates and proportions*. New York: John Wiley & Sons, Inc.
- Fridja, N.H., Ortony, A., Sonnemans, J., & Clore, G.L. (1992). The complexity of intensity: Issues concerning the structure of emotion intensity. In M.S. Clark (Ed.), *Emotion* (pp. 60–89). Newbury Park, CA: Sage.
- Greenberg, L.S., & Goldman, R. (1988). Training in experiential psychotherapy. *Journal of Consulting and Clinical Psychology*, 56, 696–702.
- Greenberg, L.S., Rice, L., & Elliott, R. (1993). *Facilitating emotional change*. New York: Guilford.
- Greenberg, L.S., & Safran, J.D. (1987). *Emotion in psychotherapy*. New York: Guilford.
- Greenberg, L.S. & Safran, J.D. (1989). Emotion in psychotherapy. *American Psychologist*, 44(1), 19–29.
- Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, 6, 278–296.
- Ivey, A.E. (1971). *Microcounseling*. Springfield, IL: Thomas.
- Ivey, A.E. (1983). *Intentional interviewing and counseling*. Monterey, CA: Brooks/Cole.
- Johnson, W.F., Emde, R.N., Scherer, K.R., & Klinnert, M.D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry*, 43, 280–283.
- Kirk, R. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Pacific Grove, CA: Brooks/Cole.
- Klein, M.H., Mathieu, P., Gendlin, E.T., & Kiesler, D.J. (1970). *The experiencing scale: A research and training manual* (Vols. 1–2). Madison, WI: Wisconsin Psychiatric Institute, Bureau of audio visual instruction.
- Koegel, R.L., & Egel, A.L. (1979). Motivating autistic children. *Journal of Abnormal Psychology*, 88, 418–426.
- Lane, R. (1991). *LEAS scoring manual and glossary*. Tucson, AZ: University of Arizona Health Sciences Center.
- Lane, R.D., Quinlan, D.M., Schwartz, G.E., Walker, P.A., & Zeitlin, S.B. (1990). The levels of emotional awareness scale: A cognitive-developmental measure of emotion. *Journal of Personality Assessment*, 55, 124–134.
- Mahoney, M.J. (1991). *Human change processes: The scientific foundations of psychotherapy*. New York: Basic Books.
- Mahrer, A.R., Stalikas, A., Boissoneault, M., Trainor, K., & Pilloud, L. (1990). A scale for assessing degree of strength of client feeling. *Canadian Journal of Counseling*, 24, 107–116.

- Robins, L.N., Helzer, J.E., Croughan, J., & Ratcliff, K.S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics, and validity. *Archives of General Psychiatry*, 38, 381–389.
- Rogers, P.L., Scherer, K.R., & Rosenthal, R. (1971). Content filtering human speech: A simple electronic system. *Behavior Research Methods and Instrumentation*, 3, 16–18.
- Rosenthal, R., Hall, J., DiMatteo, M.R., Rogers, P.L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore, MD: The Johns Hopkins University Press.
- Rosner, R. (1996). *The relationship between emotional expression, treatment, and outcome in psychotherapy*. Frankfurt: Lang.
- Scherer, K. (1981). Vocal indicators of stress. In J.K. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 171–187). New York: Grune & Stratton.
- Scherer, K.R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143–165.
- Scherer, K.R., Ladd, D.R., & Silverman, K.E. (1983). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America*, 76, 1346–1356.
- Schreibman, L., Kaneko, W.M., & Koegel, R.L. (1991). Positive affect of parents of autistic children: A comparison across two teaching techniques. *Behavior Therapy*, 22, 479–490.
- Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, 52, 1061–1086.
- Shoham-Salomon, V., Avner, R. & Neeman, K. (1989). You're changed if you do and changed if you don't: Mechanisms underlying paradoxical interventions. *Journal of Consulting and Clinical Psychology*, 57, 590–598.
- Snow, R. E. (1991). Aptitude treatment interactions as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology*, 59, 205–216.
- Wagner, H.L., Mac Donald, C.J., & Manstead, S.R. (1986). Communication of individual emotions by spontaneous facial expressions. *Journal of Personality and Social Psychology*, 50, 737–743.